HOMEWORK 2

1. Prove that the OLS estimator $\hat{\beta}$ is the same as the maximum likelihood estimator.

极大似然估计

在线性回归中, 我们假设误差项ϵ服从均值为零, 方差为σ^2的正态分布, 即$\epsilon \sim N(0, \sigma^2)$.

则线性回归模型的似然函数为:

$$L(\beta, \sigma^2 | y, X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - X_i\beta)^2}{2\sigma^2}}$$

对该表达式取对数, 得到对数似然函数:

$$\begin{aligned}
\ln L(\beta, \sigma^2 | y, X) &= \sum_{i=1}^{n} \ln \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - X_i\beta)^2}{2\sigma^2}} \\
&= -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - X_i\beta)^2
\end{aligned}$$

求一阶导数并令其为零, 我们得到:

$$\begin{aligned}
\frac{\partial \ln L(\beta, \sigma^2 | y, X)}{\partial \beta} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} X_i(y_i - X_i\beta) \\
&= \frac{1}{\sigma^2} (X^\top y - X^\top X\beta) = 0
\end{aligned}$$

易得其正规方程为:

$$X^\top y = X^\top X\beta$$

解出β, 我们得到:

$$\hat{\beta} = (X^\top X)^{-1} X^\top y$$

OLS

易得, 其在数学上可以写成:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - X_i\beta)^2$$

同样的, 对其求导并令其为零, 我们得到:

$$\begin{aligned} \frac{\partial \sum_{i=1}^{n} (y_i - X_i\beta)^2}{\partial \beta} &= -2 \sum_{i=1}^{n} X_i(y_i - X_i\beta) \\ &= -2(X^\top y - X^\top X\beta) = 0 \end{aligned}$$

易得其正规方程为:

$$X^\top y = X^\top X\beta$$

解出β, 我们得到:

$$\hat{\beta} = (X^\top X)^{-1} X^\top y$$

总结

由上述过程可知, OLS估计量$\hat{\beta}$与最大似然估计量相同.

2. Prove the Gauss-Markov Theorem.

首先, 我们定义一个线性无偏估计量$\tilde{\beta}$为:

$$\tilde{\beta} = Ay$$

其中, A是一个$k \times n$的矩阵. 为了使$\tilde{\beta}$无偏, 我们必须有:

$$
\begin{aligned}
E(\tilde{\beta}) &= E(Ay) \\
&= AE(y) \\
&= AE(X\beta + \epsilon) \\
&= AX\beta + AE(\epsilon) \\
&= AX\beta + 0 \\
&= AX\beta
\end{aligned}
$$

等于β. 因此, 我们得到:

$$AX = I_k$$

其次, 我们计算$\tilde{\beta}$的方差:

$$
\begin{aligned}
Var(\tilde{\beta}) &= Var(Ay) \\
&= AVar(y)A^\top \\
&= AVar(X\beta + \epsilon)A^\top \\
&= AVar(\epsilon)A^\top \\
&= A(\sigma^2 I_n)A^\top \\
&= \sigma^2 AA^\top
\end{aligned}
$$

最后, 我们比较$\tilde{\beta}$和$\hat{\beta}$的方差. 根据高斯消元法, 我们可以将矩阵A分解为:

$$A = (X^\top X)^{-1}X^\top + D$$

其中, D是一个$k \times n$的矩阵, 满足:

$$DX = 0$$

将上式代入$\tilde{\beta}$的方差中, 我们得到:

$$
\begin{aligned}
Var(\tilde{\beta}) &= \sigma^2(AA^\top) \\
&= \sigma^2((X^\top X)^{-1}X^\top + D)((X^\top X)^{-1}X^\top + D)^\top \\
&= \sigma^2((X^\top X)^{-1}X^\top + D)(X(X^\top X)^{-1} + D^\top) \\
&= \sigma^2((X^\top X)^{-1} + DD^\top)
\end{aligned}
$$

由于$DD^\top$是一个半正定矩阵, 即对任意非零向量v, 都有$v^\top DD^\top v \geq 0$, 因此我们有:

$$Var(\tilde{\beta}) - Var(\hat{\beta}) = \sigma^2 DD^\top \geq 0$$

即$\tilde{\beta}$的方差不小于$\hat{\beta}$的方差.

高斯-马尔可夫定理得证.

3. Prove $E(\hat{\sigma}^2) = \sigma^2$

要证明$E(\hat{\sigma}^2) = \sigma^2$，我们需要假设误差项$\epsilon$服从均值为零，方差为$\sigma^2$的正态分布，即$\epsilon \sim N(0, \sigma^2)$。这意味着因变量$y$也服从均值为$X\beta$，方差为$\sigma^2$的正态分布，即$y \sim N(X\beta, \sigma^2)$。

给定X和β时，y的似然函数为：

$$L(y|X, \beta) = \prod_{i=1}^{n} f(y_i|x_i, \beta)$$

其中$f(y_i|x_i, \beta)$是正态分布的概率密度函数在y_i处的取值。代入正态密度的公式，我们得到：

$$L(y|X, \beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right)$$

对两边取自然对数，我们得到对数似然函数：

$$\log L(y|X, \beta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2$$

要找到σ^2的最大似然估计量，我们需要对σ^2最大化对数似然函数。求一阶导数并令其为零，我们得到：

$$\frac{\partial \log L(y|X, \beta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 = 0$$

解出σ^2，我们得到：

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2$$

这与OLS估计量$\hat{\sigma}^2$完全相同。因此，我们证明了$E(\hat{\sigma}^2) = \sigma^2$。

4.

Given conditions:

> **(A1) The relationship between response (y) and covariates (X) is linear;**
>
> **(A2) X is a non-stochastic matrix and rank(X) = p;**
>
> **(A3) $E(\epsilon) = 0$. This implies $E(y) = X\beta$;**
>
> **(A4) $cov(\epsilon) = E(\epsilon\epsilon^\top) = \theta^2 I_N$; (Homoscedasticity)**
>
> **(A5) ϵ follows multivariate normal distribution $N(0, \epsilon^2 I_N)$ (Normality)**

Prove the following results:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^\top X)^{-1}) \tag{0.1}$$

$$(N-p)\hat{\sigma}^2 \sim \sigma^2 {X_{N-p}}^2 \tag{0.2}$$

(0.1)

由于, $\hat{\beta}$ 是一个线型函数 ϵ 的函数, 而 ϵ 服从正态分布，所以根据正态分布的性质，$\hat{\beta}$ 也服从正态分布。我们只需要计算 $\hat{\beta}$ 的均值和方差即可。

由于 $E(\epsilon) = 0$ 所以:

$$\begin{aligned}
E(\hat{\beta}) &= E((X^\top X)^{-1} X^\top y) \\
&= E((X^\top X)^{-1} X^\top (X\beta + \epsilon)) \\
&= E((X^\top X)^{-1} X^\top X\beta + (X^\top X)^{-1} X^\top \epsilon) \\
&= E(\beta + (X^\top X)^{-1} X^\top \epsilon) \\
&= \beta + (X^\top X)^{-1} X^\top E(\epsilon) \\
&= \beta + 0 \\
&= \beta
\end{aligned}$$

由于 $cov(\epsilon) = \sigma^2 I_N$, 所以:

$$\begin{aligned}
cov(\hat{\beta}) &= cov((X^\top X)^{-1} X^\top y) \\
&= cov((X^\top X)^{-1} X^\top (X\beta + \epsilon)) \\
&= cov((X^\top X)^{-1} X^\top X\beta + (X^\top X)^{-1} X^\top \epsilon) \\
&= cov((X^\top X)^{-1} X^\top \epsilon) \\
&= (X^\top X)^{-1} X^\top cov(\epsilon) X (X^\top X)^{-1} \\
&= (X^\top X)^{-1} X^\top \sigma^2 I_N X (X^\top X)^{-1} \\
&= \sigma^2 (X^\top X)^{-1}
\end{aligned}$$

所以 $\hat{\beta} \sim N(\beta, \sigma^2 (X^\top X)^{-1})$.

(0.2)

由于 $\hat{\sigma}^2$ 是一个ϵ的二次函数，而 ϵ 服从正态分布，所以根据卡方分布的性质，$\hat{\sigma}^2$ 也服从卡方分布:

由 $E(\epsilon) = 0$:

$$
\begin{aligned}
E((N-p)\hat{\sigma}^2) &= E(\sum_{i=1}^{N}(y_i-\hat{y}_i)^2) \\
&= E(\sum_{i=1}^{N}(X\beta+\epsilon_i-X\hat{\beta})^2) \\
&= E(\sum_{i=1}^{N}(X(\beta-\hat{\beta})+\epsilon_i)^2) \\
&= E(\sum_{i=1}^{N}\epsilon_i^2+(\beta-\hat{\beta})^T(X^T\epsilon+\epsilon^T\hat{\beta})) \\
&= E(\sum_{i=1}^{N}\epsilon_i^2+0) \\
&= E(tr(\epsilon^T\epsilon)) \\
&= tr(E(\epsilon^T\epsilon)) \\
&= tr(cov(\epsilon)) \\
&= tr(\sigma^2 I_N) \\
&= N\sigma^2
\end{aligned}
$$

由于 $cov(\epsilon)=\sigma^2 I_N$，所以 ϵ 的每个分量都互相独立，并且都服从正态分布，所以根据卡方分布的性质，$(N-p)\hat{\sigma}^2$ 的自由度等于 ϵ 的秩，即 N 减去 X 的秩，即 $N-p$。

综上所述，我们可以得出结论：$(N-p)\hat{\sigma}^2 \sim \sigma^2 {X_{N-p}}^2$

5. Suppose y follows the log-linear regression relationship with $x \in (\mathbb{R})^p$, i.e.,

$$log(y) = x^\top \beta + \epsilon \tag{0.3}$$

where ϵ follows normal distribution $N(0, \sigma^2)$. Please calculate $E(y)$.

由于 $log(y) = x^\top \beta + \epsilon$，而 $\epsilon \sim N(0, \sigma^2)$，所以把 $log(y)$ 看作是一个服从正态分布的随机变量, 由于$y = e^{log(y)} = e^{x^\top \beta + \epsilon}$，而 ϵ 和 $x^\top \beta$ 是相互独立的随机变量（因为 ϵ 和 x 都是非随机的），所以我们可以把 y 看作是两个相互独立的随机变量的乘积，即 $y = e^{x^\top \beta} e^{\epsilon}$, 易得:

$$\begin{aligned} E(y) &= E(e^{x^\top \beta} e^{\epsilon}) \\ &= E(e^{x^\top \beta}) E(e^{\epsilon}) \\ &= e^{x^\top \beta} e^{E(\epsilon) + \frac{1}{2} Var(\epsilon)} \\ &= e^{x^\top \beta + 0 + \frac{1}{2}\sigma^2} \\ &= e^{x^\top \beta + \frac{1}{2}\sigma^2} \end{aligned}$$

综上:

$$E(y) = e^{x^\top \beta + \frac{1}{2}\sigma^2}$$

6. Define $\hat{y}_i = {x_i}^\top \beta$. Let the intercept be included in the regression model. Define the total sum of squares (TSS) and explained sum of squares (ESS) as follows

$$TSS = \sum_i (y_i - \overline{y})^2$$
$$ESS = \sum_i (\hat{y}_i - \overline{y})^2$$

Please prove:

$$TSS = ESS + RSS$$

$$\begin{aligned} TSS &= \sum (y_i - \overline{y})^2 \\ &= \sum ((y_i - \hat{y}_i) + (\hat{y}_i - \overline{y}))^2 \\ &= \sum ((y_i - \hat{y}_i)^2 + (\hat{y}_i - \overline{y})^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \overline{y})) \\ &= RSS + ess + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \overline{y}) \end{aligned}$$

所以:

$$
\begin{aligned}
\sum(y_i-\hat{y}_i)(\hat{y}_i-\overline{y}) &= \sum(y_i-\hat{\beta}_0-\hat{\beta}_1x_i)(\hat{\beta}_0+\hat{\beta}_1x_1-\overline{y}) \\
&= \sum(y_i-\overline{y}+\hat{\beta}_1\overline{x}-\hat{\beta}_1x_i)(\overline{y}-\hat{\beta}_0+\hat{\beta}_1x-\overline{y}) \\
&= \sum((y_i-\overline{y})-\hat{\beta}_1(x_1-\overline{x}))\hat{\beta}_1(x_i-\overline{x}) \\
&= \sum(\hat{\beta}_1(x_i-\overline{x})(y_i-\overline{y})-\hat{\beta}_1{}^2(x_1-\overline{x})^2) \\
&= \hat{\beta}_1\sum(x_i-\overline{x})(y_i-\overline{y})-\hat{\beta}_1{}^2\sum(x_1-\overline{x})^2
\end{aligned}
$$

由于:

$$
\hat{\beta}_1=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2}
$$

所以, 易得:

$$
\begin{aligned}
\sum(y_i-\hat{y}_i)(\hat{y}_i-\overline{y}) &= \frac{[\sum(x_i-\overline{x})(y_i-\overline{y}]^2)}{\sum(x_i-\overline{x})^2}-\frac{[\sum(x_i-\overline{x})(y_i-\overline{y})]^2\cdot\sum(x_i-\overline{x})^2}{[\sum(x_i-\overline{x})^2]^2} \\
&= \frac{[\sum(x_i-\overline{x})(y_i-\overline{y}]^2)}{\sum(x_i-\overline{x})^2}-\frac{[\sum(x_i-\overline{x})(y_i-\overline{y}]^2)}{\sum(x_i-\overline{x})^2} \\
&= 0
\end{aligned}
$$

由此, 可证:

$$
\begin{aligned}
TSS &= RSS+ESS+0 \\
&= RSS+ESS
\end{aligned}
$$